

Posudek vedoucího diplomové práce

Jméno a příjmení autora posudku: Doc. RNDr. Irena Holubová, Ph.D.

Jméno a příjmení autora práce: Bc. Tomáš Novella

Název práce: Web Data Extraction

Cílem práce bylo nejprve zanalyzovat oblast nástrojů pro extrakci strukturovaných informací z nestrukturovaných dat na webu. Na základě analýzy a zjištěných nedostatků stávajících přístupů pak autor navrhl, implementoval a také úspěšně prakticky aplikoval vlastní řešení – jazyk Serrano – s důrazem na rozšiřitelnost, modularitu a využití v prostředích s omezeními.

Hlavní klady práce:

- Autor provedl přehlednou analýzu existujících metod extrakce dat z webu, včetně různých přístupů k jejich klasifikaci.
- Navržená metoda je jasně zasazená do kontextu aktuálních přístupů, jejich klasifikací i odborných diskusí požadavků na extrakci dat. Autor diskutuje zvolené konstrukty a řešení kriticky srovnává s existujícími přístupy.
- Jazyk je navržen s důrazem na využití stávajících osvědčených webových technologií (JSON, Javascript, DOM, ...).
- Navržený přístup byl úspěšně aplikován v praxi v rámci několika projektů.

Text práce je jasný a přehledný, zahrnuje veškeré očekávané součásti. Jazyková úroveň je velmi dobrá. Sazba je profesionální.

Připomínky k práci:

- Navržený přístup by si zasloužil rozsáhlejší praktické srovnání s podobnými přístupy (což je poměrně netriviální úkol).
- Navržený jazyk by bylo možné popsat formálněji. Nicméně stávající popis nebrání pochopení.

Přes výše uvedené připomínky práce bezpochyby splňuje původní zadání i cíl. Student dobře pronikl do zvolené problematiky a prokázal schopnost detailně zanalyzovat a řešit zadaný problém. Práce Tomáše Novelly podle mého názoru splňuje podmínky kladené na diplomovou práci.

Doporučení k obhajobě:

Z výše uvedených důvodů práci *doporučuji* k obhajobě.

V Praze dne: 16. 8. 2016

Podpis: